

# *elexiko* – GESCHICHTE UND ZUKUNFT

Ulrike Haß ist Professorin für Linguistik der deutschen Sprache an der Universität Duisburg-Essen und leitete von 1999-2005 die Abteilung Lexik am Institut für Deutsche Sprache in Mannheim.

Annette Klosa ist seit 2002 wissenschaftliche Mitarbeiterin der Abteilung Lexik am Institut für Deutsche Sprache in Mannheim.

Im Juni 2004 wurden der Fachöffentlichkeit die ersten knapp 250 lexikografisch bearbeiteten Stichwörter im Onlinewörterbuch *elexiko* präsentiert; gut zehn Jahre später wurde die Bearbeitung des „Lexikons zum öffentlichen Sprachgebrauch“ (das Kernmodul in *elexiko*) abgeschlossen und bei einem Kolloquium im Februar 2015 zur Diskussion gestellt. Dies war auch Anlass, auf die Anfänge des Projekts zurückzublicken, Bilanz über das Erreichte zu ziehen und Gedanken für die Zukunft von *elexiko* zu formulieren.

Von diesen Geschichten vom Anfang handelt dieser Beitrag. Er beschreibt außerdem die Veränderungen der Benutzeroberfläche und den fortwährenden Ausbau des Wörterbuches um vielfältige (redaktionell erarbeitete und automatisch gewonnene) Angaben. Als Forschungsprojekt hatte bzw. hat *elexiko* nicht nur die Aufgabe, ein Wortschatz-Informationssystem zu konzipieren und zu realisieren, sondern alle damit verbundenen (lexikologischen, metalexikografischen, korpus- und computerlinguistischen) Aspekte zu untersuchen. Deshalb werden hier auch einige Zahlen zu den Forschungsergebnissen rund um das Wörterbuchprojekt und die Partizipation des Projektteams an der Fachdiskussion bei Vorträgen, Tagungen, in Netzwerken usw. präsentiert. Schließlich geben wir noch einen Ausblick auf die weitere Pflege der erarbeiteten Substanz sowie die Fortsetzung der Arbeit in neuen Modulprojekten.



Abb. 2: Sprachforschung am IDS in den 1960er Jahren: Deutsch im Zettelkasten (li. Ulrike Hauser-Suida, re. Gabriele Hoppe)

## Geschichten vom Anfang

### Vorgeschichten und Ausgangssituation

Hätten die Lexikografen des 16., 17., 18. und 19. Jahrhunderts Datenbanken gekannt, hätte man sie nicht überreden müssen, ihre hölzernen Ordnungshilfen, die Zettelkästen, gegen elektronische einzutauschen, und sie hätten auch kaum gezögert, den Nutzern einen besseren Zugang zum Wortschatz-Wissen als denjenigen über das alphabetische Buch vorzuschlagen. Die ‚natürliche‘ Form der Wortschatzbeschreibung ist nicht das Buch, sondern die Datenbank, und der Zettelkasten (vgl. Abb. 1) war die Übergangsform.



Abb. 1: Zettelkästen des Projekts Deutsches Fremdwörterbuch

Jacob Grimm verglich sein Wörterbuch einführend mit einem Wäscheschrank und seinen Fächern; er präsentiert das Wörterbuch also als eine sehr ordentliche und saubere Angelegenheit ‚für Frauen‘:

**frauen**, mit ihrem gesunden mutterwitz und im gedächtnis gute sprüche bewahrend, tragen oft wahre begierde ihr unverdorbnos sprachgefühl zu üben, vor die **kisten und kasten** zu treten, aus denen wie gefaltete leinwand **lautere wörter** ihnen entgegen quellen: ein wort, ein reim führt dann auf andere und sie kehren öfter zurück und **heben den deckel** von neuem. (Jacob Grimm, Vorrede DWB 1854, Spalte XIII, Hervorh. UH)

Mitarbeiterinnen des IDS waren neben ihren männlichen Kollegen bereits in der prädigitalen Ära lexikografisch tätig (vgl. Abb. 2).

```

131.00      MAIN-ANF
132.00      PERFORM INIT-01
133.00      IF KCRCCC > "000"
134.00          PERFORM FEHLER
135.00          PERFORM MPUT-FECHO1
136.00          PERFORM PEND-FI
137.00      ELSE
138.00          IF KCTACAL = "ECHOTAC"
139.00              MOVE SPACES TO FECHO1
140.00              MOVE ECHOTAC TO MELDUNG1
141.00              PERFORM MPUT-FECHO1
142.00              PERFORM PEND-FI
143.00          ELSE
144.00              PERFORM MGET-FECHO1
145.00              evaluate true
146.00              when kcrccc = "30a"
147.00                  if kchanid = "admin" or = "chef"
148.00                      perform admin-ende
149.00                      perform pend-fi
150.00                  else
151.00                      move spaces to facho0
152.00                      perform mput-facho0
153.00      STOP HOST OR CONNECTION CURRENTLY INOPERABLE
154.00
155.00      0131.00:001(0)
156.00
157.00      HOLL SAN DUE
158.00
159.00      TAST FEHL

```

Abb. 3: BS 2000

Weil also der ‚elektronische Zettelkasten‘, d. h. die Datenbank, als ‚natürliche‘ Speicher- und Präsentationsform des Wörterbuchs gelten kann, lag es nahe, dass das computeraffine IDS früher als irgendein germanistisches Universitätsinstitut auf den Gedanken kam, sich der datenbankbasierten Lexikografie zuzuwenden.

## ANFÄNGE DER DATENBANKBASIERTEN LEXIKOGRAFIE

Den Anfang machte 1982 die Konzeption LEDA (Lexikographische Datenbank; Brückner 1982). LEDA war eine konzeptionelle Vorstufe und führte noch nicht unmittelbar zu einem lexikografischen Projekt. Wesentlich war aber schon bei LEDA die enge Verbindung zwischen der Idee einer lexikografischen Datenbank und den Mannheimer Textkorpora, die damals schon eine beträchtliche Größe aufwiesen. REFER war die von Tobias Brückner entwickelte Recherchesoftware, Vorgänger des heutigen COSMAS II. Mit REFER ‚auf‘ den IDS-Korpora zu arbeiten, hieß in den 1980er Jahren, auf einen Monitor mit grüner oder bernsteingelber Schrift auf schwarzem Grund zu schauen, Befehle einzutippen, denn es gab noch keine Maus, und seine Ausdrücke auf großformatigem, gelochtem Endlospapier im Erdgeschoss des IDS abzuholen, wo der sog. Kettendrucker stand, der genauso viel Lärm machte, wie die Bezeichnung nahelegt. Das Betriebssystem hieß damals BS 2000, stammte von Siemens und war der „neueste Schrei“ (vgl. Abb. 3). Das REFER-Benutzerhandbuch von 1989 enthält noch Abschnitte wie: Woraus besteht ein Computer? Was ist eine Datei? Da musste man als Linguistin und Linguist ‚durch‘.

Der nächste Schritt hin zu *lexiko* / OWID war das in den 1980er Jahren im IDS neu konzipierte Wörterbuch der „Brisanten Wörter“. Es wurde zwar noch in Form von Text auf elektrischen Speicher-Schreibmaschinen getippt, basierte aber schon wesentlich auf den Korpo-

ra, die zu diesem Zweck auch um domänenspezifische und von den Lexikografen selbst ausgewählte Texte erweitert wurden; diese Texte sind heute im sogenannten Handbuchkorpus im DeReKo enthalten.

## GRAMMIS UND LEXIS

1993 startete *Grammis* (ab 2000 **grammis**), damals noch mit dem Namenszusatz „Die Grammatik mit der Maus“ (vgl. Abb. 4). Der Untertitel signalisierte einerseits technische Modernität, andererseits enthielt er eine Anspielung auf die damals bereits 20 Jahre alte Kinderfernsehsendung des WDR, die für freches Bildungsfernsehen stand – eine Assoziation, die den **grammis**-Kollegen gefallen haben dürfte.



Abb. 4: *Grammis*-Logo von 1993

Jedenfalls sollte, was **grammis** für die Abteilung Grammatik war, nämlich eine Art Zentral- oder Vorzeigeprojekt, auch für die Abteilung Lexik geschaffen werden. Damit war keinesfalls eine wissenschaftliche Bewertung verbunden, sondern folgte einem durch das Aufkommen des Internets verstärkten Trend, dem zufolge man in der Öffentlichkeit für Forschung werben und bestmöglich sichtbar werden musste. Das neue Projekt sollte daher die öffentliche Sichtbarkeit einer Abteilung stärken, die zwar lange schon mit diversen Wörterbuchprojekten aktiv war, aber noch nicht



Abb. 5: Das *elexiko*-Logo im Design von 2007

im Internet. Lexikografie sollte sich nach außen sichtbar in die digitale Welt begeben. Hinzu kam, dass die Abteilung Lexik Anfang der 1990er Jahre personell um Kolleginnen und Kollegen aus der Berliner Akademie angewachsen war, die Ideen und ‚Arbeitswut‘ mitbrachten.

Logisch, dass das Schwesterprojekt von **grammis** „Lexis“ heißen musste. Es stellte sich aber rasch heraus, dass es eine amerikanische Firma gab, die zudem als eine der ersten mit Volltextrecherchen und Online-Informationsangeboten Geld verdienen wollte: „lexisnexis“, wobei „lexis“ eine selbständige Komponente war, sodass tatsächlich eine Namensgleichheit bestanden hätte. Auch eine Änderung der Schreibung in „Lexxis“ oder „Leksis“ verringerte das juristische Risiko nicht. Der schöne Name fürs Projekt war also unhaltbar, und so musste ein neuer Name her. Eine deutsche Bezeichnung bot sich an, die deutlich machte, dass hier etwas anderes als ein Wörterbuch entstehen sollte. ‚Wissen‘ galt und gilt immer noch als ein ‚buzzword‘ und steht, im Unterschied zu ‚Information‘, für einen Mehrwert. Das Problem war die Abkürzung, „WüW“ schied selbstverständlich aus. Daher wählten wir „WiW“, das eine semantisch willkommene Homophonie mit einem französischen Wort (‘vif’, ‘vive’) hat. Dennoch fand „WiW“ mittelfristig wenig Anklang, so dass im Laufe der konzeptionellen Arbeit das Kunstwort *elexiko* gebildet wurde, mit dem wir dann auch online gingen (wie in Abb. 5 gezeigt, wurde die Phrase „Wissen über Wörter“ aber noch lange als Untertitel weitergeführt). Die sachliche Unterscheidung zwischen OWID und *elexiko* kam erst 2008, als die Portalfunktion und die Fülle der damit verbundenen Aufgaben von den eigentlichen lexikografischen Aufgaben getrennt wurden.

### **elexikos Motivationen**

Es gab mehrere Gründe, warum das IDS Anfang der 1990er Jahre ein digitales Informationssystem zum deutschen Wortschatz realisieren wollte. Ein digitales Wörterbuch im Sinne eines Informationssystems bot die Chance, mehrere laufende Wörterbücher und auch neue lexikologische Projekte einerseits weiterzuführen, andererseits zwischen ihnen Bezüge herzustellen. Daher war von Beginn an klar, dass eine sehr flexible Struktur benötigt würde, um die Untersuchungs-

schwerpunkte aller, auch neuer Projekte, abzudecken. Gewissermaßen als Rückgrat der Vielfalt sollte eine umfangreiche Stichwortliste dienen, die alle Stichwortmengen der Einzelprojekte in sich einschloss. 300.000 Wörter wurden von den Kollegen aus der Korpuslinguistik nach bestimmten Kriterien ermittelt.

Dies alles konnte nur modular, Stück für Stück geschehen, und für Modularität im Aufbau sprach neben dem Genannten noch ein ganz anderer, praktischer Grund: Um 300.000 Wörter zu bearbeiten, hätte man auch bei effizientester Arbeitsweise gigantisch viele Personennjahre benötigt. Und es dachte auch niemand daran, dem Duden Konkurrenz zu machen, der in dieser Zeit die digitalen Medien ebenfalls für sich zu nutzen begann. *elexiko* sollte etwas ganz anderes sein als die kleine Schwester des Duden. Es sollte exemplarische Ausschnitte aus dem Gesamtwortschatz sprachwissenschaftlich fundiert beschreiben, die Möglichkeiten der hypertextuellen Präsentation ausnutzen und dabei allgemeinverständlich sein. Dies sollte zunächst an einem Demonstrationswortschatz ausprobiert und der Fachwelt vorgestellt werden.

Als genuin hypertextuelles Wörterbuch wollte *elexiko* sich grundlegend von allen lexikografischen Digitalisaten unterscheiden, die damals v. a. als CD-ROM auf den Markt kamen und die das Aussehen des ‚Wörterbuchs‘ mehr oder weniger bewahrten. *elexiko* wollte die Artikelstruktur von vornherein als nicht-linear entwerfen und traditionelle lexikografische Gewohnheiten notfalls auch radikal infrage stellen.

### **MIT HYPERTEXT EINE BREITE ANGEBOTSPALETTE SCHAFFEN**

Gemessen mit heutigen Maßstäben war der Einfluss potenzieller Nutzer auf die Anlage von *elexiko* noch gering, was auch an der noch unentwickelten Nutzerforschung lag. Es blieb uns nichts anderes übrig, als verschiedenste Rezeptionspfade gleichermaßen anzulegen und eine Angebotspalette zu schaffen, auf der man in jeder Situation etwas Passendes findet.

Während wir heute gewohnt sind, von Usern und Usability zu sprechen, wenn es um Anpassung an Wörter-

buchadressaten geht, waren damals andere Kategorien wichtig. Es waren eher kollektive Kategorien wie ‚interessierte Öffentlichkeit‘ oder ‚demokratisch-aufgeklärte Gesellschaft‘, deren sprachreflexiven und sprachkritischen Interessen ein neues Wörterbuch unabhängig vom Medium entgegenkommen sollte. Dass Forschungsprojekte auch extern repräsentiert werden und dass sie sich legitimieren sollen, hat das IDS schon sehr früh akzeptieren und umsetzen müssen. Für *ellexiko* hieß das von vornherein, dass es sowohl wissenschaftliche Nutzerinnen und Nutzer als auch interessierte Laien adressieren musste. Ein rein wissenschaftlichen Zwecken dienendes Informationssystem war undenkbar. Ein ungelöstes Problem blieb freilich lange Zeit die mit dem modularen Prinzip einhergehende unvollständige Abdeckung der alphabetischen Stichwortstrecke.

## DIE NUTZER VON *ellexiko*

Hieraus erklärt sich auch die Wahl des ersten Wortschatzmoduls „Lexikon zum öffentlichen Sprachgebrauch“. Dieser frequente Wortschatz wurde u. E. in keinem anderen Wörterbuch angemessen beschrieben und setzte ein Korpus wie das Mannheimer geradezu voraus. *ellexiko* sollte also strikt korpusbasiert erarbeitet werden. Introspektiv gewonnene Angaben waren tabu und die mit der Zeit immer interessanter werdenden Möglichkeiten der Kollokationsanalyse faszinierend. Hinzu kam der zeitlich parallele Aufbau von DeReKo in Kooperation mit den Universitäten Tübingen und Stuttgart. Die Korpusbasis, die *ellexiko* nutzen konnte, wurde immer besser.

Gespalten waren die Meinungen innerhalb des Instituts für Deutsche Sprache über der Frage, ob automatisch erzeugte Angaben, v. a. Beispielbelege, gut genug seien oder nicht. Sie hätten viel Manpower und Geld gespart. So gut wie die Verfahren heute sind, waren sie damals und lange Zeit später noch nicht. Und auch heute sind sie ja nicht immer gut genug.

## *ellexikos* Anfänge

Zwischen den ersten konzeptionellen Überlegungen und der Online-Veröffentlichung des 250 Lemmata umfassenden Demonstrationswortschatzes im Jahr 2003 verging eine relativ lange Zeit. Das lag an der Situation der Software. Für beinahe beliebige Anwendungen gibt es heute Open-Source-Software, die nicht nur kaum etwas kostet, sondern auch oft viel niedrigere Hürden für Anwender aufwirft. Was wir damals suchten, war eine XML-fähige objektrelationale Datenbank, und die gab es noch nicht.

Die Strukturierungssprache XML kam 1998 auf, also kurz nach der Geburt der Idee von „Lexis“, und es war klar, dass für uns nur XML infrage kam, um die Artikelstruktur für unser Internet-„Wörterbuch“ zu definieren. Das Schwesterprojekt *grammis* nutzte die Datenbank Oracle, aber die am IDS verfügbare Version von Oracle war damals noch nicht XML-fähig. Daher versuchten wir es zunächst mit dem Dokumentenmanagementsystem „Tamino“ der Software-AG, Darmstadt. Aber es stellte sich nach etlichen Workshops und Gesprächen heraus, dass dieser Tamino weder zaubern noch flöten konnte und vor Wörterbuchstrukturen kapitulierte.

Während wir noch mit „Tamino“ kämpften, während wir uns selber und mit Carolin Müller-Spitzers Hilfe ein Stück weit XML beibrachten, um die Denkweise zu verstehen und auf ein Wörterbuch abzubilden, wurde Oracle dann XML-fähig: Es konnte also endlich losgehen. 2003 ging *ellexiko* mit dem Demonstrationswortschatz online und stellte sich der Öffentlichkeit und der Fachdiskussion.

## WÖRTERBÜCHER IN DER DISKUSSION

*ellexiko* zeichnet sich seither durch eine besondere Artikelstruktur aus, die einerseits dem ‚state of the art‘ im Sinne der Heidelberger Wörterbuchforschung entsprach, und die andererseits die Fesseln der Linearität abwarf und einen Wörterbuchartikel als hypertextuellen Raum wahr machte. Der Einfluss der Heidelberger Wörterbuchforschung war wichtig, weil schon die räumliche Nähe für einen langjährigen und intensiven





Abb. 6: Startseite von *elexiko*  
im Dezember 2003

Austausch der Ideen sorgte. Seit den 1980er Jahren nahmen IDS-Mitarbeiter an Herbert Ernst Wiegands Heidelberger Lexikografischem Kolloquium teil, genauso wie die Mitglieder der Duden-Redaktion. Daraus entstand die Olms-Reihe „Wörterbücher in der Diskussion“, aber der regelmäßige und über Jahre hinweg geführte, persönliche Diskurs ist noch einmal etwas anderes als eine umfassende Lektüre.

Glücklicherweise aber drehten sich diese Diskussionen noch kaum um die Konsequenzen der Digitalisierung oder gar Hypertextualisierung von Wörterbüchern. Es gab noch freien Gedankenraum für uns, und die Diskussionen der Projektgruppe um Lemmatypen, obligatorische wie fakultative Angaben sowie ihre Anordnung gehören zu den konstruktivsten Diskussionen, die man sich vorstellen kann. Wesentliche Ergebnisse dieser Diskussionen zeigen in *elexiko* / OWID heute ihre Nachhaltigkeit: etwa, dass der Zusammenhang der Lesarten erläutert wird, d. h., dass semantische Variation thematisiert wird, dass Kollokationen und Syntagmen nicht nur einfach gelistet, sondern quasi unter Bedeutungsaspekten (framesemantisch) eingeordnet werden, und dass Besonderheiten des Gebrauchs nicht nur punktuell, sondern systematisch berücksichtigt werden.

### PRAGMALINGUISTIK, KOGNITIONSLINGUISTIK UND KORPUSLINGUISTIK

Die darin erkennbare sprachtheoretische Basis war zunächst eine pragmalinguistische. Es ging um die Beschreibung des Gebrauchs, nicht um semantische Merkmale oder dergleichen. Einerseits war dies in den 1990er Jahren selbstverständlich, andererseits war es lexikografisch relativ neu, denn alle großen wissenschaftlichen Wörterbücher basierten wegen ihrer langen Laufzeit immer noch auf strukturellen Vorstellun-

gen – auch das „Wörterbuch der deutschen Gegenwartssprache“ (WDG), das sich in elektronifizierter Form damals gerade zum Konkurrenten von *elexiko* entwickelte. Die pragmalinguistische Basis wurde aber um kognitionslinguistische Einsichten erweitert, für die sich besonders Claudia Fraas (heute TU Chemnitz) einsetzte, die sich aber auch ein Stück weit aus den Korpusdaten – corpus driven – ergab. Das IDS-Korpus war inzwischen gigantisch groß, und es war uns allen sehr bewusst, dass es für *elexiko* weit mehr und etwas ganz anderes als ein Beleglieferant war. Das Korpus begann mit uns zu sprechen: „Was sagt denn das Korpus?“ oder „Frag doch mal das Korpus!“ waren Standardsätze der Projektdiskussionen. Den Zusammenhang zwischen Kognition und Korpus theoretisch und grundsätzlich zu klären, hat dann später eine Essener Dissertation (Wrede 2013) unternommen.

Zu Beginn wurde nicht zwischen *elexiko* als Rahmen gebendem Informationssystem und den *elexiko*-Modulen ‚Demonstrationswortschatz‘ und ‚Öffentlicher Sprachgebrauch‘ unterschieden, wie es heute die Differenzierung zwischen OWID und *elexiko* tut. Andere Module waren von Anfang an vorgesehen und in Vorbereitung, insofern waren für die festen Mehrwortverbindungen, die Kommunikationsverben oder den Schuldiskurs – um nur einige zu nennen – eine spezifische Angabestruktur und ein eigenes Design zu entwerfen. Dies leistete ja die ‚Auszeichnungssprache‘ XML und bot uns die Möglichkeit, die Verknüpfung aller lexikologischen Projekte mit der Sichtbarmachung ihrer Besonderheiten und ‚Alleinstellungsmerkmale‘ in Einklang zu bringen.

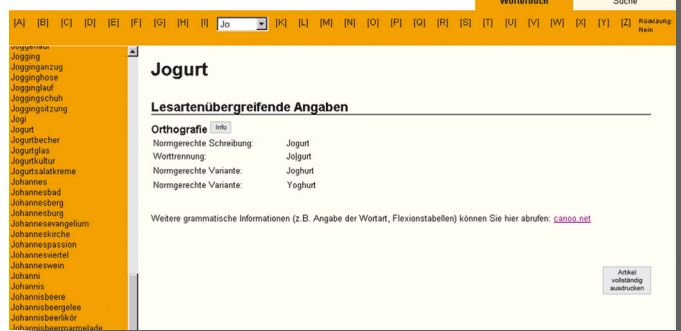


Abb. 7: Orthografische Angaben zu den *elexiko*-Stichwörtern (Dezember 2003)

Obwohl es nach außen schwer zu vermitteln war, dass *elexiko* kein schlichtes Nachschlagewerk für 300.000 Lemmata sein sollte, war die Alternative, die exemplarische Darstellung einiger Wortschatzbereiche, nie infrage gestellt. Warum? Weil ein lexikalisches Informationssystem, das an einem Forschungsinstitut entsteht, im Unterschied zu einem digitalen Wörterbuch nicht primär auf die Befriedigung von Nutzerbedürfnissen ausgerichtet, sondern eine medial innovative Form der Darstellung wortschatzbezogener Linguistik sein sollte. Das schließt nicht aus, dass die interessierten Laien auch etwas davon haben. Aber die Priorität lag bei der Forschungsleistung und Innovation.

## INNOVATIVE DARSTELLUNG WORTSCHATZBEZOGENER LINGUISTIK

Es war beim Entwurf der Artikelstruktur z. B. sehr wichtig, die Korrelation wissenschaftlicher Kategorien untersuchen zu können und erstmals Abfragen zu ermöglichen, die z. B. den Grad der Polysemie mit der Wortbildungsart oder mit der Korpusfrequenz verknüpfen oder die semantische Klasse mit irgendeinem anderen Merkmal in Beziehung setzen zu können. Was uns vorschwebte, war eine Art Data Mining über dem gefüllten Informationssystem. Zwar wuchsen die Bäume dann in der Realisierungsphase weniger hoch in den Himmel, aber immerhin steht einiges davon heute in den „Erweiterten Suchen“ der in OWID gebündelten Lexikon-Projekte zur Verfügung.

Die „Kisten und Kasten voller lauterer Wörter“, von denen Jacob Grimm sprach, werden in der noch unbekannten medialen Zukunft immer wieder neu interpretiert werden. Forschungsinteressen und Kenntnisse

ändern sich. Eigentlich sind sowohl Kästen als auch Datenbank-Strukturen zu starr für das, was Sprecher als Wort bezeichnen. Zettelkasten, Wörterbuch und Datenbank – alle sperren sie die Wörter ein und halten lediglich Momentaufnahmen ihres kommunikativen Lebens fest. Aber die Kästen, ob digitale oder aus Papier, sorgen doch dafür, dass Wörter den Fragenden und Betrachtern nicht weglaufen, sondern sich aus der Nähe betrachten lassen.

## Der weitere Projektverlauf

### Die Entwicklung von *elexiko* im Überblick

Wie ging es nach den Vorgeschichten und Anfängen mit *elexiko* weiter? Wie konnte die linguistisch-lexikografische Konzeption in den Folgejahren umgesetzt werden, und wie hat sich die Benutzeroberfläche von *elexiko* entwickelt? Diesen Fragen möchten wir im Folgenden nachgehen.

Im Dezember 2003 ging *elexiko* unter der URL <www.elexiko.de> online. Im Angebot waren damals nur die Stichwortliste mit den orthografischen Angaben (z. B. Schreibvarianten sowie Angaben zur Worttrennung) sowie einige Wörterbuchaußentexte, in denen das Projekt und einige Suchmöglichkeiten vorgestellt wurden (zur Startseite des damaligen Onlineauftritts vgl. Abb. 6, zur Ansicht der Angaben zu einem Stichwort vgl. Abb. 7).

Erste Recherchemöglichkeiten nach Stichwörtern mit bestimmten Merkmalen rundeten das Angebot ab (damals allerdings beschränkt auf bestimmte orthografische Merkmale, vgl. Abb. 8). Im Suche-Menü konnten außerdem einfache Wortsuchen und Operatorensuchen durchgeführt werden. Der Zugriff auf die Wortartikel war schließlich (so wie noch heute) auch über die Buchstabenleiste möglich.

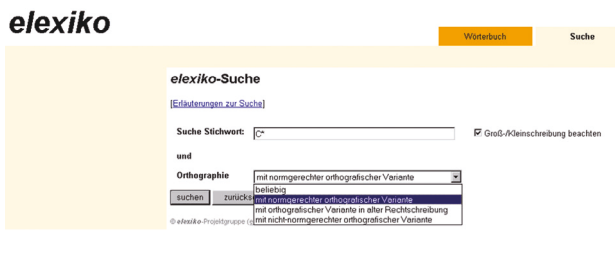


Abb. 8: Die Suchfunktion in *ellexiko* im Dezember 2003

Das Onlineangebot orientierte sich in Aufbau und Design einerseits an der damaligen IDS-Webseite sowie an *grammis*, dem grammatischen Informationssystem des IDS, was auch in dessen heutigem Design noch spürbar ist (vgl. Abb. 9). In diesem zweispaltigen Layout wurden in der linken Spalte die Stichwortliste bzw. die Menüstruktur der Wörterbuchaußentexte angezeigt, und der größere rechte Bildschirmbereich wurde zur Anzeige der Wörterbuchaußentexte bzw. der Angaben zu den Stichwörtern genutzt.

Im Juni 2004 wurden die ersten redaktionell bearbeiteten Stichwörter des „*ellexiko*-Demonstrationswortschatzes“ online publiziert (und parallel in einem Kolloquium am IDS der Fachöffentlichkeit vorgestellt). Im Folgejahr begann das Projekt mit der Erarbeitung des „Lexikons zum öffentlichen Sprachgebrauch“, das Neologismenwörterbuch der 1990er Jahre (verantwortlich Doris Steffens) wurde in das *ellexiko*-Portal integriert und der von Ulrike Haß herausgegebene Band „Grundfragen der elektronischen Lexikographie. *ellexiko* – das Online-Informationssystem zum deutschen Wortschatz“, der die lexikologisch-lexikografische Konzeption zu *ellexiko* enthält, erschien.

Für die Darstellung redaktionell bearbeiteter Stichwörter wurde anfangs der rechte, inzwischen der mittlere Bildschirmbereich genutzt, wobei immer zunächst lesartenübergreifende Angaben und außerdem die Navigationsstruktur, über die die verschiedenen Lesarten geöffnet werden können, präsentiert wird (vgl. Abb. 10). Nach Auswahl einer Lesart werden die hierauf bezogenen Angaben in einem neuen Bildschirm präsentiert (vgl. Abb. 11). Anfangs nutzte das Projekt noch verschiedenfarbige Hintergründe und Rahmen, um bestimmte Angaben hervorzuheben (z. B. Paraphrase, Belege, Kommentare und Hinweise), was im neuen OWID-Design aufgegeben wurde. Benutzerhinweise (aufrufbar durch Button) bzw. Info-Knopf-Texte wurden bzw. werden z. B. in Pop-up-Fenstern angezeigt.

In den folgenden beiden Jahren konnte zunächst das Onlineangebot des Projekts „Usuelle Wortverbindungen“ (2006, verantwortlich: Kathrin Steyer), dann das „Wörterbuch zum Schuldiskurs 1945-1955“ (2007, verantwortlich: Heidrun Kämper) ins Portal integriert werden. Im Jahr 2007 wurde das *ellexiko*-Portal auch in ein neues Design gegossen, das bei einem Kolloquium zum Thema „Wörterbuchportale“ am IDS präsentiert wurde.<sup>1</sup> Einen Eindruck von dieser Entwicklungsphase gibt Abbildung 12, in der die mittlerweile stark ausgebaut Suchfunktion nach (lexikografisch bearbeiteten) Stichwörtern mit bestimmten Merkmalen gezeigt wird.



Abb. 9: Das „Grammatische Wörterbuch“ in *grammis* (Februar 2015)

Mit der Einwerbung eines Drittmittelprojekts zum Thema „Benutzeradaptive Zugänge und Vernetzungen in *ellexiko* (BZVelexiko)“ aus den Mitteln des Paktes für Forschung und Innovation des Senatsausschusses Wettbewerb (SAW) der Leibniz-Gemeinschaft (Antragstellerinnen: Carolin Müller-Spitzer und Annette Klosa) begann 2009 eine dreijährige Phase, in der das Projekt *ellexiko* durch die Arbeiten im Drittmittelprojekt flankiert und gefördert wurde. Im Rahmen dieses

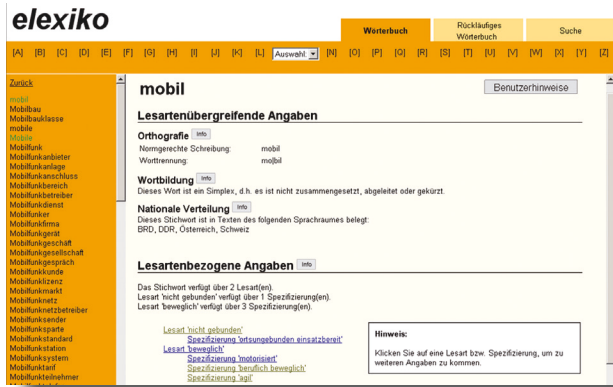


Abb. 10: Erste Bildschirmseite bei bearbeiteten *exlexiko*-Stichwörtern (Juni 2004)



Abb. 11: Lesartenbezogene Angaben bei bearbeiteten *exlexiko*-Stichwörtern (Juni 2004)

Projekts wurden Benutzungsstudien zum Wörterbuch durchgeführt, ein Tool, mit dem Vernetzungen z. B. zwischen paradigmatischen Partnern angelegt und verwaltet werden können, konnte entwickelt werden, und Angaben zu Wortbildungsprodukten konnten aus einer Datenbank mit morphologischen Angaben entnommen werden.

2010 hat das Projekt die praktischen Erfahrungen bei der korpusgestützten Erarbeitung des Wörterbuchs in einem weiteren Sammelband „*exlexiko*. Erfahrungsberichte aus der lexikographischen Praxis eines Internetwörterbuchs“ (herausgegeben von Annette Klosa) dokumentiert. Während der Ausbau des „Lexikons zum öffentlichen Sprachgebrauch“ voranschritt und auch die Erweiterung des automatisch gewonnenen Angabespektrums (z. B. automatisch ausgewählte Korpusbelege) vorangetrieben werden konnte, wurde im Projekt OWID eine Neukonzeption des Portals (sowohl hinsichtlich der Struktur wie des Designs) entwickelt, die 2011 der Öffentlichkeit vorgestellt wurde. Auch das Wörterbuch *exlexiko* ist seitdem im OWID-Design online, wie in Abbildung 13 anhand der ersten Bildschirmseite zum Stichwort *Behörde* gezeigt. Geändert hat sich hier u. a., dass sich das Portal erstmals in einem dreispaltigen Layout präsentiert, wobei die rechte Spalte für die Menüpunkte des jeweiligen Wörterbuchs genutzt wird (vgl. Abb. 13).

Die Benutzeroberfläche von *exlexiko* konnte 2012 aufgrund einiger der in den o. g. Benutzungsstudien erhobenen Daten weiter optimiert werden. Mit dem Abschluss der Arbeiten am „Lexikon zum öffentlichen Sprachgebrauch“ publizierte das Projekt 2014 schließ-

lich ein Themenheft der Zeitschrift „Deutsche Sprache“, in dem u. a. verschiedene Benutzungsmöglichkeiten der lexikografisch erarbeiteten Wortartikel vorgestellt werden.

### *exlexiko* in Zahlen<sup>2</sup>

An *exlexiko* haben bis heute 19 Mitarbeiterinnen und Mitarbeiter und 20 studentische (geprüfte und ungeprüfte) Hilfskräfte mitgearbeitet. Außerdem haben 20 Praktikantinnen und Praktikanten von 12 verschiedenen Hochschulen insgesamt 113 Wochen (bzw. gute zwei Jahre) im Projekt verbracht. Alle zusammen haben eine Fülle an Ergebnissen erarbeitet, die im Folgenden anhand von einigen Zahlen illustriert werden sollen. Zu den Projektergebnissen zählen dabei natürlich insbesondere die lexikografischen Informationen in <www.exlexiko.de>, aber auch die Resultate aus meta-lexikografischer und lexikologischer Forschung im Projekt, die bei Tagungen und in Publikationen der Fachöffentlichkeit präsentiert wurden.

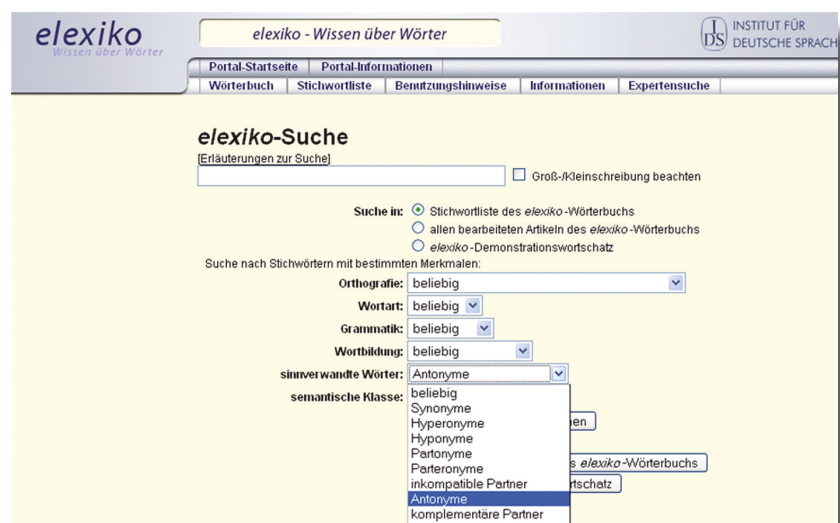


Abb. 12: Design des *exlexiko*-Portals von 2007 am Beispiel der Suchmöglichkeiten



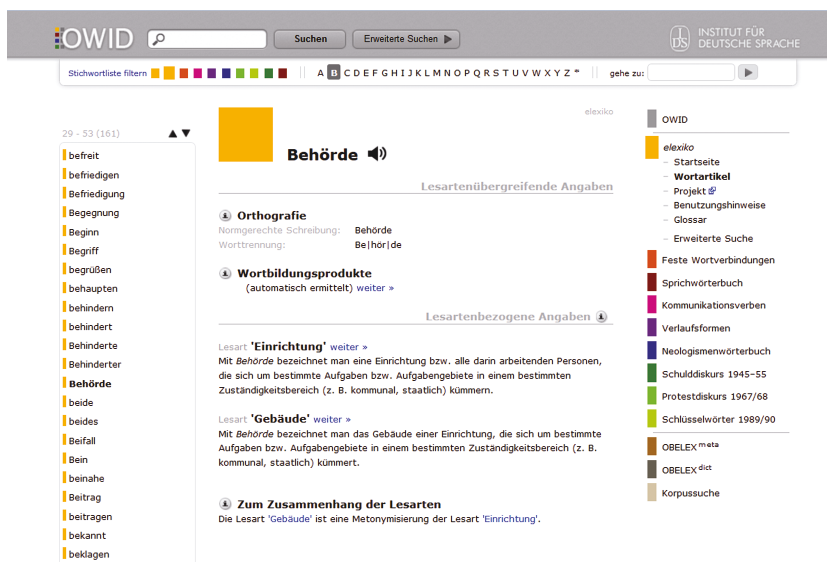


Abb. 13: *lexiko*-Wörterbuch im OWID-Design von 2011

Zunächst zu den lexikografischen Ergebnissen: Das „Lexikon zum öffentlichen Sprachgebrauch“ enthält 1.873 bearbeitete Stichwörter, von denen 870 monosem und 1.003 polysem sind. Insgesamt umfassen die bearbeiteten Stichwörter 4.129 verschiedene Lesarten (Einzelbedeutungen), d. h., durchschnittlich sind 2,21 Lesarten pro Stichwort bearbeitet und mit dem vollen Angabeinventar (Bedeutungserläuterung, Kollokationen, typische Konstruktionen, sinnverwandte Wörter, Gebrauchsbesonderheiten, Grammatik) versehen. In den 1.870 bearbeiteten Stichwörtern sind im Einzelnen enthalten:

- Angaben zu Kollokatoren: 142.946,<sup>3</sup>
- Angaben zu Konstruktionen: 44.281,
- Angaben von paradigmatischen Relationspartnern: 41.829,
- Angaben zu Gebrauchsbesonderheiten: über 1.500 (wobei über die Hälfte Angaben zu diskursgebundenen Vorkommen sind).

Außerdem enthalten die bearbeiteten Stichwörter 59.797 Belege aus dem *lexiko*-Korpus, d. h. durchschnittlich knapp 32 Belege pro Stichwort bzw. durchschnittlich etwa 14,5 Belege pro Lesart.

## DAS „LEXIKON ZUM ÖFFENTLICHEN SPRACHGEBRAUCH“

Zu 408 Lesarten gibt es im Ganzen 696 Illustrationen; insgesamt sind 239 bearbeitete Stichwörter in *lexiko* illustriert, was einem Anteil von etwa 13% entspricht. Neben den lexikografischen Angaben enthalten die bearbeiteten Stichwörter 5.399 lexikografische Kommen-

tare und Hinweise, sowie 1.104 Hyperlinks zu anderen Onlinewörterbüchern. Bei insgesamt 1.530 Stichwörtern können jeweils bis zu drei Hörbelege aus der „Datenbank für gesprochenes Deutsch“ des IDS aufgerufen werden. Die Außentexte des Wörterbuchs umfassen knapp 140 HTML-Dokumente mit Benutzungshinweisen und Glossareinträgen, 20 HTML-Dokumente mit Informationen zum Projekt *lexiko*.

Die *lexiko*-Stichwortliste umfasst etwas über 300.000 Einträge, zu denen 13.100 mal orthografische Varianten und 2.873 mal morphologische Varianten verzeichnet sind. Die nicht lexikografisch bearbeiteten (eher niedrig frequenten) Stichwörter enthalten neben orthografischen Angaben jeweils drei automatisch aus dem *lexiko*-Korpus ausgewählte Belege sowie Angaben zur Verteilung im *lexiko*-Korpus und bieten Verlinkungen zu <Canoo.net> (für grammatische Angaben) sowie zur Kookkurrenzdatenbank CCDB des IDS.

Bei 11.473 (bearbeiteten und nicht bearbeiteten) Stichwörtern gibt es schließlich Angaben zur Wortbildungsproduktivität, indem insgesamt 266.981 Wortbildungsprodukte zu simplizischen Basiswörtern angezeigt werden.

Die metalexikografischen, lexikologischen oder auch praktischen Projektergebnisse wurden bei 61 Tagungen an 27 Universitäten in 14 verschiedenen Ländern und in insgesamt 137 Vorträgen präsentiert. Dabei wurden (bei einer angenommenen Redezeit von jeweils



Abb. 14: Einige zentrale Publikationen aus dem Projekt *elexiko*

30 Minuten) insgesamt fast 70 Stunden (oder knapp drei Tage) Vortragszeit verbraucht und 114.676 km (oder etwa 2,86 Reisen um die Welt) zurückgelegt. Bei 14 Jahren bzw. 168 Monaten Projektlaufzeit bedeutet dies, dass durchschnittlich etwa alle eineinhalb Monate ein Vortrag rund um *elexiko* gehalten wurde.

### ***elexiko* IM WISSENSCHAFTLICHEN KONTEXT**

Das Projekt hat fünf Monografien, sechs Sammelbände und 111 Aufsätze (in 13 verschiedenen Fachzeitschriften) mit einer Gesamtmenge von 4.258 Druckseiten publiziert (zu einigen wichtigen Publikationen vgl. Abb. 14). Im Projektkontext sind vier Masterarbeiten und drei Dissertationen entstanden. Außerdem hat das Projekt 14 Kolloquien und Workshops veranstaltet, bei denen insgesamt geschätzte 600 Teilnehmerinnen und Teilnehmer zu Gast waren. Bei 14 Jahren Projektlaufzeit bedeutet dies, dass im Rahmen des Projekts durchschnittlich etwa 25 Druckseiten pro Monat veröffentlicht und durchschnittlich jedes Jahr ein Kolloquium (mit)organisiert wurden.

Weiterhin hat das Projekt viele Gäste (andere Lexikografinnen und Lexikografen, Linguistinnen und Linguisten, Korpuslinguistinnen und Korpuslinguisten, Sprachtechnologinnen und Sprachtechnolog) zu Besuch gehabt bzw. hat Besuche bei Kolleginnen und Kollegen in Deutschland und Europa gemacht. Das Projektteam hat schließlich zahlreiche Sprachanfragen beantwortet, sprachliche Gutachten geschrieben, Lehrveranstaltungen abgehalten und studentische Besuchergruppen oder Schülergruppen und einmal sogar eine Gruppe von Kindergartenkindern über seine Arbeit informiert.

Es bleibt festzuhalten: *elexiko* war und ist ein sehr lebendiges Projekt, an dem nicht nur viele Menschen inner- und außerhalb des Projektes mitgearbeitet haben, sondern das sich auch in vielerlei Weise an der Fachdiskussion um die korpusgestützte Erarbeitung eines modernen Wortschatzinformationssystems beteiligt hat. Es gibt wohl kaum ein Wörterbuchprojekt, das jeden Aspekt seiner Arbeit so ausführlich und nach außen hin sichtbar dokumentiert hat wie dieses. Und es gibt nur wenige Projekte, die sich so konsequent mit anderen Wörterbuchredaktionen und Forschern vernetzt und der Diskussion gestellt haben wie *elexiko*.

### **Und wie geht es weiter?**

Mit dem Projekt *elexiko* geht es selbstverständlich weiter. Der erarbeitete Bestand wird gepflegt und bei Bedarf korrigiert, außerdem soll an der automatischen Ermittlung lexikografischer Angaben, z. B. zur Gebildetheit der nicht simplizischen Stichwörter oder zur Wortart, weitergearbeitet werden. Die Recherchemöglichkeiten werden ausgebaut, und der Bestand an Artikeln wird durch regelmäßige Freischaltung von neu erarbeiteten, sogenannten Wortgruppenartikeln ergänzt. In diesen werden zwei oder mehr redaktionell bearbeitete Stichwörter in bestimmten Lesarten vergleichend beschrieben, wobei etwa Gemeinsamkeiten und Unterschiede bei den jeweiligen Kollokationen, Konstruktionen und sinnverwandten Wörtern erläutert werden. Es werden dabei drei Typen von Wortgruppenartikeln unterschieden: solche, die sinnverwandte Wörter zusammenfassen (vgl. Abb. 15 für ein Beispiel), solche, die Stichwörter eines bestimmten Themenfeldes betrachten, und Wortfelder.

Daneben wird die Arbeit beispielsweise im Projekt „Paronymwörterbuch: Ein Online-Wörterbuch zum Gebrauch verwechselbarer Ausdrücke“ fortgesetzt (vgl. Storjohann 2014). In diesem Projekt werden Ausdrücke untersucht, die formal ähnlich sind (oft gleiche morphologische Wurzel, unterschiedliche Wortbildungsmorpheme) und die über semantische Gemeinsamkeiten verfügen, und die daher von Sprecherinnen und Sprechern leicht verwechselt werden. Ziel dieses Projekts ist die linguistische und lexikografische Neubetrachtung der Paronymie, indem sie korpusanalytisch erforscht und theoretisch begründet wird, aber auch in einem elektronischen Nachschlagewerk mit kontrastiven Einträgen (innerhalb von OWID) dokumentiert wird.

## Literatur

- Brückner, Tobias (1982): Der interaktive Zugriff auf die Textdatei der Lexikographischen Datenbank (LEDA). In: Sprache und Datenverarbeitung, Heft: 1-2 / 6. Jahrgang, S. 28-35.
- Grimm, Jacob (1854): Vorrede, in: Deutsches Wörterbuch, Bd. 1, Sp. I-LXVIII. Leipzig.
- Storjohann, Petra (2014): Das künftige Projekt „*elexiko*: Paronymwörterbuch“. In: SPRACHREPORT. Informationen und Meinungen zur deutschen Sprache 30.2, S. 22-29.
- Wrede, Julia (2013): Bedingungen, Prozesse und Effekte der Bedeutungskonstruktion. Der sprachliche Ausdruck in der Kotextualisierung. Duisburg.

## Anmerkungen

- <sup>1</sup> Ein von Annette Klosa herausgegebener Band „Lexikografische Portale im Internet“ mit den Tagungsergebnissen ist 2008 erschienen.
- <sup>2</sup> Für die Unterstützung bei der Ermittlung der Zahlen danken wir Frank Michaelis (OWID) und Martin Loder (*elexiko*).
- <sup>3</sup> Bei den Kollokatoren, Konstruktionen und paradigmatischen Partnern wurden die Tokens und nicht die Types gezählt – d. h. beispielsweise, dass sich in den über 140.000 Fällen, in denen das DTD-Element für Kollokatoren gefüllt wurde, die einzelnen Kollokatoren durchaus wiederholen können.

## Bildnachweise

- Abb. S. 10, 19: IDS
- Abb. 3: <[www.stefanbion.de/dosprogs/](http://www.stefanbion.de/dosprogs/)> (14.2.2015)
- Abb. 4: <[http://hypermedia.ids-mannheim.de/call/public/fragen.ansicht?v\\_id=4519](http://hypermedia.ids-mannheim.de/call/public/fragen.ansicht?v_id=4519)> (18.3.2015)
- Abb. 9: <[http://hypermedia.ids-mannheim.de/call/public/gramwb.ansicht?v\\_app=g](http://hypermedia.ids-mannheim.de/call/public/gramwb.ansicht?v_app=g)> (28.7.2015)
- Abb. 13: <<http://www.owid.de/artikel/170765?module=elex&pos=13>> (28.7.2015)
- Abb. 15: <<http://www.owid.de/wb/elexiko/gruppen/maedchen-junge.html>> (28.7.2015)
- Abb. S. 12, 14, 15, 16, 17 (unten): Screenshots wurden von den Autorinnen erstellt; die Seiten sind inzwischen offline. ■



Abb. 15: Ausschnitt aus dem Wortgruppenartikel *Mädchen – Junge*